

## 音声情報処理研究支援システム - STRESS -

浮田輝彦 井上 博 中川聖一 坂井利之  
(京都大学 工学部)

## 1.はじめに

音声情報処理研究に、計算機は今や必須不可欠の道具として利用されており、研究室レベルでも各種の研究用システムが開発され、利用されている<sup>1)~3)</sup>。音声研究の分野は、音声分析・認識・合成など多岐に亘るが、これらの研究は、共通する部分やお互いに依存するところも多い。そのため、研究用システムは、これらの各分野において適用可能な基本的能力を備えていることが望ましい。

われわれは、以前開発したコンピュータネットワークによる音声研究システム<sup>4)</sup>を使用してきたが、今回その一部を拡張・整備し、音声情報処理研究支援システム(Speech Treating Research Support System, 以下STRESSという)を開発した。このシステムは、基本的な入出力・表示記録・処理機能を、利用者がコマンドの形で指示することにより、マン・マシン・インタラクティブに、種々の方法で、要求された機能を実行する能力を持っており、いろいろな問題に適用され、有用性が確認されている。

本稿では、このシステムの設計方針、特徴、構成、利用例について述べる。

## 2.設計方針及び特徴

本システムの設計に当たり、考慮した音声研究の分野は、音声分析・認識・合成であるが、これらの分野で、能率よく研究を進める上で有用な道具立てとしての汎用システムの開発を目標とした。音声データの性質や研究用シ

ステムの立場から整理すれば、次のような要求や問題点などを考慮する必要がある。

## (1) 音声データについて

## (a) 多量のデータ

時々刻々変化する音声のデジタル表現が必然的に膨大になる上、特に認識実験などの場合、多数の発話を扱うことが多い。これらに対し、高速大容量の記憶装置の利用による解決や、一連の処理手順のくり返し処理によって研究者の負荷を軽減する必要がある。

## (b) 処理の即時性・高速性

音声データの入出力ソフトウェアには、当然即時性が要求される上に、多量のデータであるがゆえに処理の高速性が要望される。これらは、特殊プロセッサなどにより解決されるが、研究面でのこの要求は比較的緩和される場合が多い。

## (c) データ内容の把握

音声波形を目で見ただけでは内容がわからないため、常に耳で聞いたり、各種の分析を行なうことにより内容の確認を行なう必要がある。

## (2) 研究用システムについて

(a) 研究の能率向上が得られること  
実験的研究の過程におけるアルゴリズムの決定に必要なデータや研究者の洞察に必要なデータが最小限の努力で得られる必要があり、研究に付随する作業の軽減化が要求される。これらは、研究者間のソフトウェアの共有・蓄積によって解決を計る。

## (b) 実験の試行が容易なこと

実験の際の処理手順の追加・変更が容易に行えると同時に、対話的に処理を進める能力がなければならぬ。

(c) その他

マン・マシン・システムとして、互いの特徴を相補的に組合せ、総合的研究能力を向上しうる機能を提供する必要がある。

われわれは、以上の要求を満たすため、音声データについては、特殊プロセッサの導入など主にハードウェア的側面から、また研究用システムについては、主にソフトウェアの側面からの支援を行う方針を採った。後者についてシステム設計時に特に留意した点は、次の二点である。

- i) 研究者間のソフトウェアの共有蓄積が容易なこと。
- ii) マン・マシン・システムとして利用できる能力を提供すること。

第一の点は、主に研究能力の向上に関連するが、音声研究の場合、研究者間で共通の処理（例えば、音声の入出力、データ収集、基本的な分析、結果の表示など）を行なう場合が少なくない。このように、他の目的のために開発されたソフトウェアの一部であっても、それを利用することが可能になると、研究のための作業が軽減され、必要な部分に努力を集中することができる。これは、いわゆるサフルーテンのライフライ化によって可能ではあるが、利用するサフルーテンの出力変数の形や主記憶上での配置など実行時の形態を考慮する必要があり、不満が残る。この点を改善するためには、コマンドにより、基本的な処理を指示するようにすれば、処理時間に若干の無駄時間が予想されるが、利用は容易になる。

第二の点は、人間と機械との間だけで、例えばポリウムなどの制御でモニターを見ながら、単に1ステップずつ処理を進めるだけでは不十分であり、

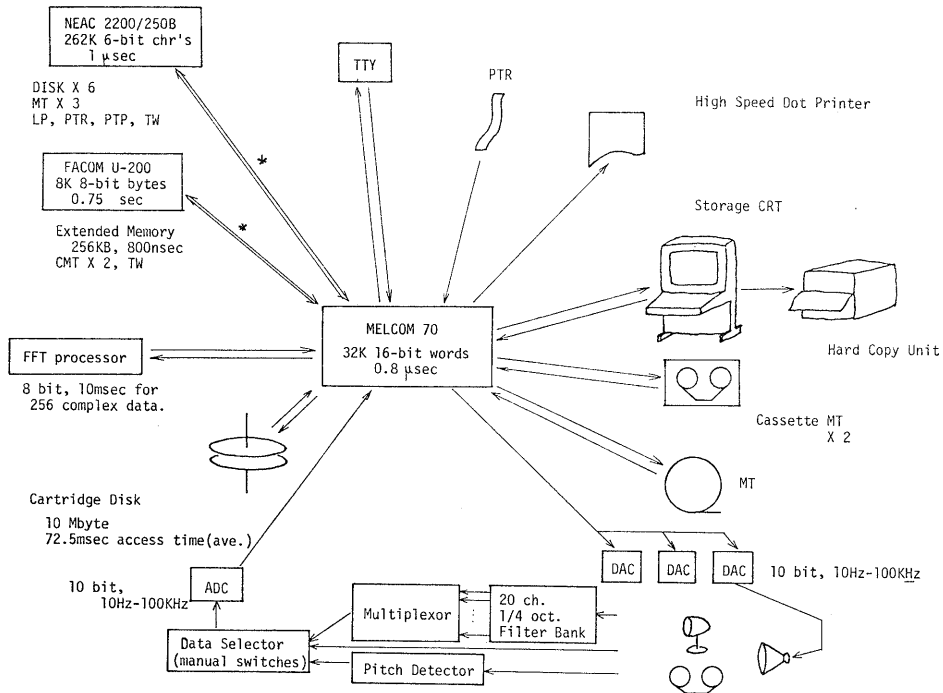


Figure 1 Hardware configuration.

\* Logical computer complex by using modified and reduced HOST-HOST protocol of the KUIPNET.

人間の監視下に機械が人手を介さず、自動的にパラメータを変化させて処理を進め、必要があれば、人間が結果を修正しうる機能が必要であることを意味している。この機能を備えることにより、不完全ではあるが、比較的簡単なアルゴリズムを利用して、多くの場合については、機械が自動的に処理し、間違った場合や例外的な場合には、人間が指示することによって、全体として能率よく処理を進め、研究者の負荷を軽減することができる。

以上のような考察や操作性及びメーカー提供のソフトウェアとの互換性、拡張性を考慮して、音声処理用の基本的機能をコマンドとして利用者に提供することにした。本システムは、次の特徴を持っている。

- (1) コマンドを単位として、音声処理に必要な基本的機能を提供する。
- (2) コマンドの処理方法として、インタラクティブ処理モードとくり返しループを含む蓄積コマンド処理モードがあり、これらのモード間の切換えが任意の時点で可能である。
- (3) コマンドパラメータに変数指定が可能である。
- (4) コマンド蓄積機能により、インタラクティブに処理して行ったコマンド系列を即時に再実行できる。
- (5) メーカー提供のOS (BDOS) を内包しているため、利用者プログラムの開発・実行用のコマンドと、音声処理用コマンドを区別せず利用できるのも、デバックが容易になる他、新コマンドの登録も簡単に行える。
- (6) コマンドの使用説明及びコマンド生成援助を行なうコマンドがある。

### 3. システム構成

#### 3.1. ハードウェア

図1にSTRESSのハードウェア構成を示す。多数の入出力機器が接続されているが、このうち一部のものについて仕様を簡単にまとめておく。A/D, D/A等については、文献[4]を参照されたい。

- (1) フィルターバンク  
20チャンネル, 1/4オクターブ,  
中心周波数 210 - 5660 Hz.
- (2) ピッチ抽出器  
3チャンネルのLPFからなるHalsey型,  
ピッチ抽出範囲 80 - 400 Hz.
- (3) FFTプロセッサ  
8ビット, 256ポイントの複素数データに対し,  
計算時間約10 msec.
- (4) カートリッジ型磁気ディスク装置  
10 Mbyte, 平均アクセス時間 = 72.5 msec,  
転送速度 = 312 Kbyte/sec.

一方、音声入出力機器が結合されており、STRESSの中核であるMelcom 70は、豊富なファイル装置を持つ汎用中型計算機Neac 2200/250Bと、256 Kbyteのバルクメモリを持つU200との間に、ネットワークを利用した論理的なコンピュータ・コンプレックスを形成しており<sup>5)</sup>、STRESSでは、プロセス間通信機能やファイル転送機能などを利用者に

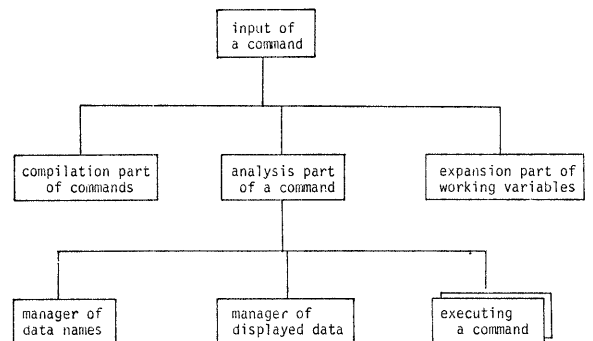


Figure 2 Software configuration.

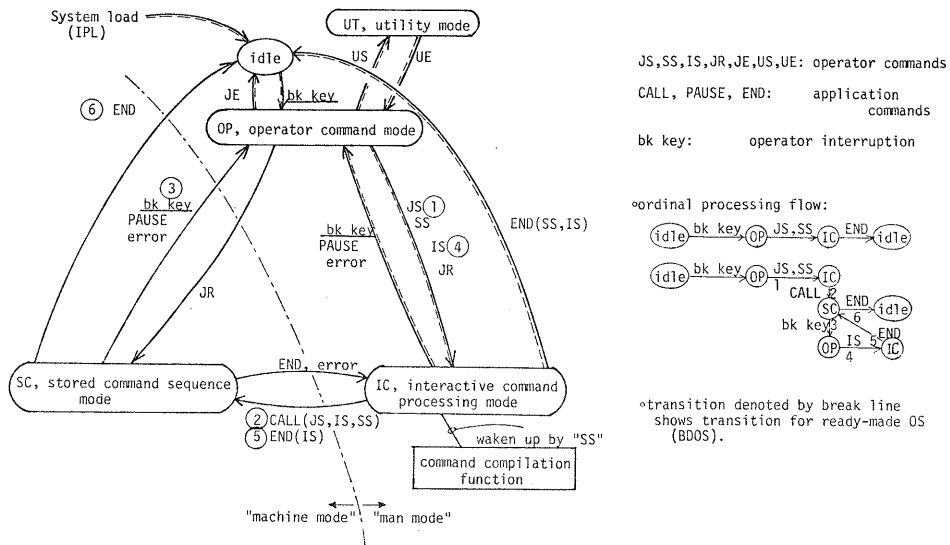


Figure 3 Commands processing modes.

提供している。

### 3.2.ソフトウェア

STRESSのソフトウェアシステムは、主に Melcom 70上で開発し、利用者に対する仕様としては、設計開始時に既に使用されていたメーカー提供のOS (BDOS) を内包する形で作成した。図2にSTRESSのソフトウェア構成の概略を示す。コマンドは、オペレータ・コマンドとアプリケーション・コマンドに分れ、システムに対し、種々の動作や処理方法を指定する。アプリケーション・コマンドの処理モードは、図3にあるように、蓄積コマンド処理モードとインタラクティブ処理モードがあり、それぞれマシンによる自動処理とマンによる対話的処理に対応しているものと考えることができる。また、これらのモード間の切換えは自由に行えるので、マン・マシン・システムとして、人間の判断力と機械の計算力とを組合せ、高水準の能力を持ったシステムを構成することができる。これは、図3

では、CALLコマンドによる蓄積コマンド系列の起動(①,②)、利用者による割込み(③)、インタラクティブ処理モードによるコマンドの実行(④)、元の蓄積コマンド系列への復帰(⑤)、というような形で実行される。

#### 3.2.1 オペレータ・コマンド

STRESSのオペレータ・コマンドには、表1のようなものがあり、アプリケーション・コマンドの処理モードの指定等を行なう。このうち、SSとIS以外は既存のBDOSと互換性を保つために設けられたものである。ここでは、SSとISについて述べる。

Table 1 List of operator commands.

* common for STRESS, BDOS	
command	
JS*	job start
JR*	job restart
JE*	job stop
US*	utility start
SS	command compilation function start
IS	interactive command processing mode start

### (1) SS (コマンド蓄積機能スタート)

このコマンドにより、コマンド蓄積機能が開始され、以下処理されるアプリケーション・コマンドの系列がシステム固有のディスク上に蓄えられるので、後に即時に再実行可能となると同時に、再入力の必要なしに、そのコマンド系列を利用し、拡張して行くことができる。

### (2) IS (インタラクティブ処理モードスタート)

このコマンドは、蓄積コマンド系列実行中に、利用者が何らかの処理を強制的に行いたい場合に用いられ、蓄積コマンド系列との会話は、作業変数(3.2.3参照)やマーク(4.1参照)を通じて行われる。

### 3.2.2. アプリケーション・コマンド

STRESS では、各種の音声処理機能をアプリケーション・コマンドとして利用者に提供している。このコマンドは BDOS のジョブ・コントロール・コマンドと全く同様に使用することができ、並にすべてのジョブ・コントロール・コマンドは STRESS のアプリケーション・コマンドとして実行可能となっている。このコマンドは、基本的に単独で実行可能でもあり、他のコマンドと組合せて一連の系列として実行することも可能である。また、蓄積コマンド系列は、既存のソースプログラムエディタによって作成・編集ができる。

表 2 に現在システムに登録されているコマンドのうち主なものの一覧表を挙げる。

### 3.2.3. 作業変数

作業変数は、アプリケーション・コマンドの制御パラメータを変数指定或いは短縮指定するために導入したもので、整数型と文字型のデータが取扱え、通常各コマンドが解釈される前に文字列として展開される。また、実行時に

Table 2 List of application commands (a part).

input/output of speech data	
RAWINPUT	input of raw speech
FILINPUT	input of filter bank output data
PLAY	playback of raw data
display and edition of speech wave	
DSPCRT	display of speech data on CRT display
CUT	cut of wave
SPLICE	splice of wave
synthesis of speech	
TANALG	synthesis of speech data by terminal analog
analysis of speech	
ANALYZE	calculation and display of speech data (eg. spectral envelopes)
working variables and control	
SET	definition of working variables
GET	input of working variable value
PUT	output of value of working variable
FNEXT	switch of file name defined for working variable
IF	check of content of working variable
CALL	wake up of stored command sequence
GOTO	Jump
END	end of job or stored command sequence
others	
REGIST	permanent registration of temporal file
DTNM	label definition of displayed data
DPNM	display of labels
SVMN	re-definition of doubly defined label
MARK	marking to displayed data
SVMK	re-definition of doubly defined mark

オペレーターが文字列の入力を要求できる特殊記号を使用することも可能で、この記号を使うと、例えば、プログラム開発の際に、アセンブル、リンク、ロードなどの手順を蓄積コマンド系列で用意しておけば、ソースプログラム名を入力するだけで、自動的に一連の処理を進めることができる。

## 4. 音声処理用ソフトウェア

各処理ソフトウェアは、各々コマンドにより動作するが、大別すると(表 2 参照)、音声データ入出力部、波形表示編集部、合成部、分析部に分けることができる。ここでは、これらについて述べ、あわせて使用例も示す。

### 4.1. データ管理

音声データは、磁気ディスク上では、エリア名とファイル名によって識別されるが、STRESS では、一度アクセス

されたデータファイルは、以後データ名(ファイル名)のみで指定することができる方法を採用している。その時、既に定義済のデータ名を定義した場合は、警告のメッセージが出力され、後に再定義することが可能となっている。

音声データ及びその一部分の指定方法には、1)データ名による方法、2)表示されている軸による方法、3)マークによる方法、の3種類があり、利用者は、求めたい場所を種々の方法で指定することができる。特にマークは、利用者がマークを指定して、ある位置につけることが可能なほか、ある処理コマンドにより、ある性質を持っている場所にマークをつけることができ、応用範囲は広い。

STRESS で取扱えるデータのタイプは、音声生データ、フィルターバンク出力のスペクトルデータ(20個/フレーム)、スペクトル包絡データ(128個/フレーム)などの時系列データが標準タイプのデータとして取扱えるとともに、利用者定義のデータについても、フ

イル名によるデータへのアクセス等の支援を行う。

一方、一時的なファイルも自由に扱うことが可能となっており、通常のデータ名と区別せず扱えるラベルをつけて処理される。これらのデータ名を整理・確認するために、データ名表示コマンドが用意されている。

#### 4.2.音声データ入出力部

入力部は、音声生データ入力と、フィルターバンクによって処理されたスペクトルデータの入力の2種類がある。また出力(再生)には、生データの出力コマンドが用意されており、データの再生による確認や、聴取実験用データの出力が行える。

#### 4.3.波形表示・編集部

表示部は、一般に蓄積型CRTディスプレイ上に時系列データを表示するもので、表示の際には、データの位置、時間スケール、振幅スケール、表示位置などが指定できる。表示データは、表示管理部によって集中的に管理されており、前記の時系列データのほかに、

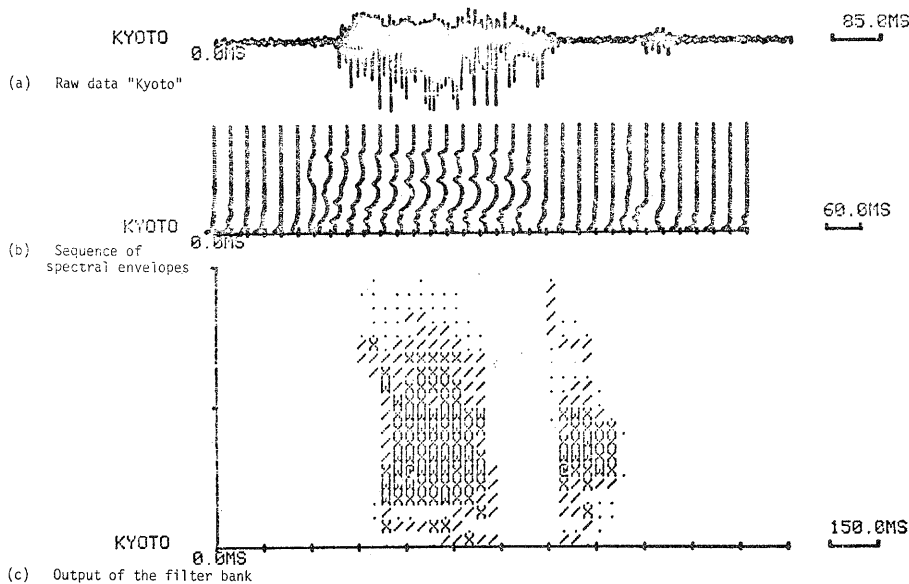


Figure 4 Examples of display command "DSPCRT".

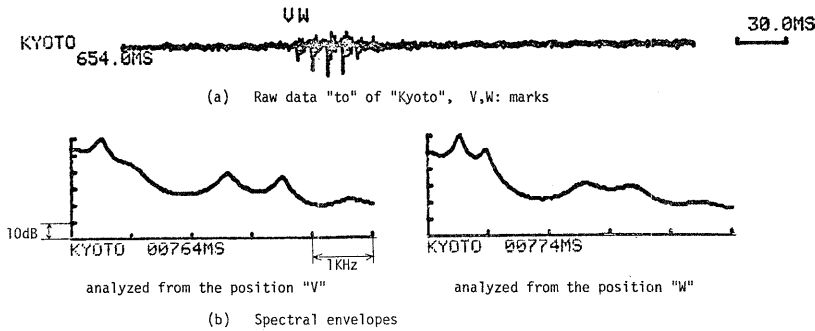


Figure 5 An example of analysis command "ANALYZ".

各種の分析結果の表示も、表示位置などが管理される。また画面が一杯になると消去するが、一画面前のデータも軸番号を指定することにより再び表示することができ、また一画面前の状態に戻すことも可能である。図4は、「京都」という発話の生データ表示(a)、LPC分析によるスペクトル包絡の表示(b)、フィルターバンクによるスペクトルデータの表示(c)である。

一方、編集部は、音声波形の切断及び結合が可能で、聴取実験用サンプルの作成や音声分析・認識に必要なデータの収集を行うことができる。またデータの切断は、表示コマンドとデータ名定義コマンドを組合せて行うことができ、あるデータの一部が表示されている時にデータ名を付与し、実際的な切断を行うことも可能である。

#### 4.4. 合成部

ターミナル・アナログ型の音声合成用シミュレータが用意されており、有声音の合成音声の作成ができる。

#### 4.5. 分析部

主にオペレータが、音声生データの内容を確認するためのもので、生データに対して、指定された位置のスペクトル表示やスペクトル包絡の表示を行う。図5は、「KYOTO」の最後の「O」の部分についてのスペクトル包絡を表示させたものである。

一方、時系列の音声生データを全時間長にわたって分析し、スペクトルやその包絡を求めるコマンドも用意されており、その際には、窓関数の指定などを行うことができる。

#### 4.6. サブルーチン・ライブラリ

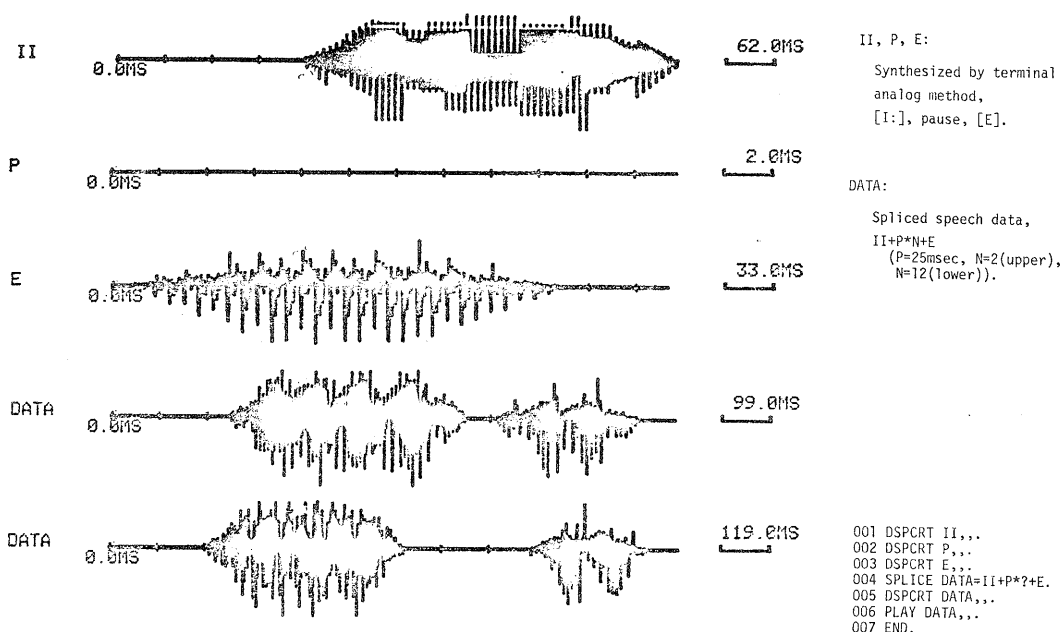
コマンドとして利用者に音声処理用ソフトウェアを提供する他、STRESSでは、サブルーチンの形でも提供しているため、利用者は各自のプログラムに最適な形で組み込み利用することができる。表3に一覧表を示すが、これらはすべてフォートラン言語とアセンブリ言語のプログラムから使用できる形に用意されている。

### 5. 利用例

本システムは、現在までに、聴取実験用サンプルの作成や、話者の個人性を表現している音響パラメータの研究、

Table 3 Subroutine library.

speech input output macro's	AD/DA macro's
softwares for analysis of command parameters	data name search, displayed data search, set values of parameters.
softwares for display and output of data	CRT display of various acoustic parameters, output to Dot Printer, hard copy.
softwares for speech analysis	digital filters, FFT, LPC, cepstrum, pitch calculation, formant extraction.
softwares for statistics	basic statistics, analysis of variance, principal component analysis.



(a) Original and spliced data

(b) Command sequence

Figure 6 An example of speech samples for hearing test.

単語認識の研究などに利用されているが、ここでは、波形編集機能の利用例として、聴取実験用サンプルの作成と、また、蓄積コマンド処理と各種分析ソフトウェアの利用例として、話者の個人性の研究の一部として行われた音声分析処理<sup>6)</sup>を示す。

### 5.1 聴取実験用サンプルの作成

本システムは、波形の編集機能を備えており、音声素片をあらかじめ用意しておき、それらをつなぎあわせることにより、聴取実験用サンプルの作成が簡単に行える。図6(a)に示すのは、「イイ」と「エ」及び無音部分を、あらかじめ、ターミナル・アナログ合成方式により用意し、それらを結合することにより、「いいえ(No)」と「良い絵」が知覚上区別される時間的要因の研究<sup>\*</sup>のために作成されたサンプルで、「イイ」と「エ」の無音区間長が、

\*京都大学文学部言語学教室の研究に、このシステムが利用された。

種々異なる<sup>サンプル</sup>を作成した(図では、DATAのうち、上が50msec、下が300msecの無音区間を挿入したものである)。同図(b)に、その時のコマンド系列を示すが、図中の「?」により、実行時に無音区間長の長さを指定できるので、このコマンド系列を呼び出すだけで、必要とするサンプルを順次作成することができる。

### 5.2 音声分析処理例

音声に含まれる個人性を表現している音響パラメータの研究の際の例で、特に多数の発話に対して、同じ処理を施し、種々の音響パラメータの抽出を行う作業が必要となる。図7(a)は、処理の結果得られるパラメータを表示したもので、同時に出力される実際の値とあわせて利用すれば、人間の勝れた洞察力により、音響パラメータの選択等に有効であった。また同図(b)は、その時のコマンド系列であり、入力フ



イル(U10 に対する作業変数 IN で、ATINDF は U10 経由でデータを入力する。)を順次切替えて、分析システムである ATINDF が自動的に実行されて行く。(この場合出力ファイルは ATINDF 固有のものを使用している。)

## 6. おわりに

音声の入出力や基本的処理機能を、コマンドとして、対話的に、また半自動的に利用できる音声情報処理研究支援システムについて述べた。現在用意されているソフトウェアは、音声のすべての研究分野について十分であるとは言い難いが、将来開発されるソフトウェアを組み込み、更に充実した研究支援システムに拡張・整備して行くための、基本的能力は充分備えていると思われる。今後、特に音声認識の成果の導入や、コンピュータ・コンプレッ

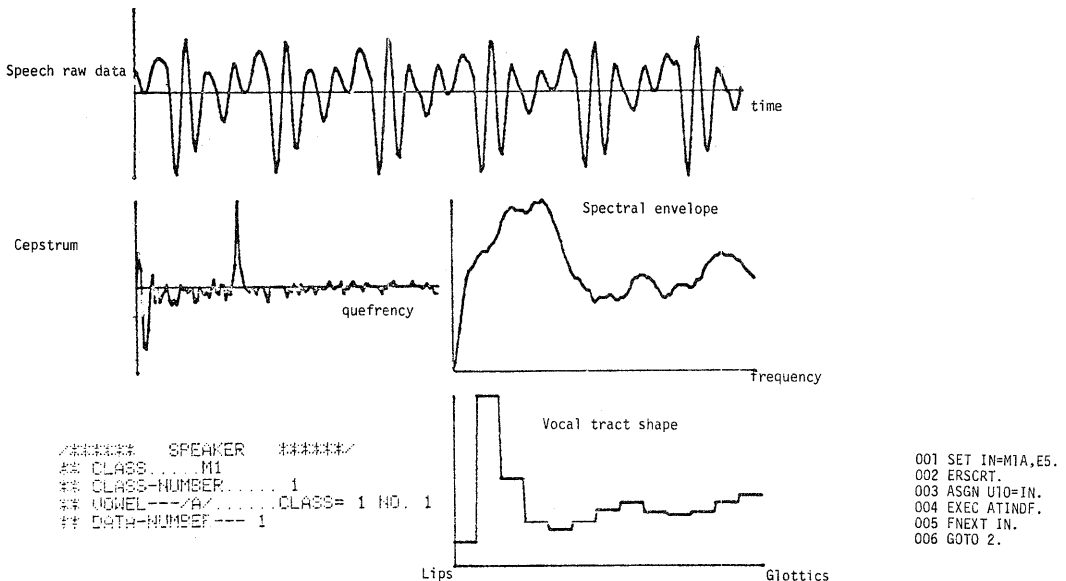
クスを利用した大規模なシステムに発展させて行きたい。

## 謝辞

システム開発にあたり、有益な助言をいただいた本学金出武雄助教授に感謝いたします。また、修士論文の一部として音声分析ソフトウェアを開発して下さった白方博教氏、波形編集部の開発を担当して下さった神田文男氏・中野潔氏、開発時に援助をいただいた石川憲洋氏・山尾雅之氏の皆様に感謝いたします。

## 参考文献

- 1) 中島隆元, 鈴木虎三, 大村浩: 音声研究用データファイル制御システム, 日本音響学会音声研究会資料, 573-07, 1973.
- 2) Nakatani, L.H.: Computer-aided signal handling for speech research, JASA, vol.61,



(a) Various acoustic parameters displayed

(b) Command sequence

Figure 7 An example of analysis of various acoustic parameters.

no.4, April, 1977.

- 3) 小林豊、新見康永：対話型音声分析システム、日本音響学会音声研究会資料、S78-43, 1978.
- 4) 坂井利之、大谷謙治：コンピュータネットワークによる音声研究システム、日本音響学会音声研究会資料、S73-32, 1973.
- 5) 会田雄一：計算機網を利用した計算機複合体の開発、京都大学修士論文、昭和53年3月.
- 6) 白方博教：母音・性別による日本語母音の特徴パラメータの変動と母音の認識、京都大学修士論文、昭和54年3月.